THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR ....Brenda Lynne Chawner........

TITLE OF THESIS ....An alternative personal retrieval system........

....

....

DEGREE FOR WHICH THESIS WAS PRESENTED ....Master of Library Science....

YEAR THIS DEGREE GRANTED ....Fall, 1978....

DATED ....October....

THE UNIVERSITY OF ALBERTA

AN ALTERNATIVE PERSONAL RETRIEVAL SYSTEM

by

© BRENDA LYNNE CHAWNER

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF LIBRARY SCIENCE

EDMONTON, ALBERTA

FALL, 1978

ABSTRACT

It is a common practice for people involved in research to take note of books, articles, reports and other materials relevant to their work. As a collection of such references grows, it becomes more and more difficult to locate specific ones, so that some form of retrieval system, however simple, is needed. This thesis examines the characteristics and use of a personal document collection, then semi-quantitatively evaluates retrieval systems presently available to a researcher. Finally, a simple alternative system, based on a computer program providing access to records stored in a sequential file, is described.

ACKNOWLEDGEMENT

This thesis would not have been possible without the help and encouragement of my supervisor, Dr. Charles Davis. I would also like to thank the National Research Council of Canada for one year of financial support.

Table of Contents

I. INTRODUCTION

As the body of literature in a researcher's field grows, it is likely that he will begin to keep a file of references to material that interests him. The problem of providing access to such a file is one that has been recognised for a number of years, and becomes more acute as the file's size increases. Several retrieval systems have been adapted for use by an individual researcher; these range from traditional author/title/subject card systems to relatively complex ones involving extensive subject analysis and subsequent coding of information.[1] More recently, computer-based systems have been described.[2]

This thesis will first examine the nature and use of personal collections, based on a survey of the literature. I will then analyse the traditional manual retrieval systems, as used by an individual researcher. Next, the application of modern data base management systems, such as SPIRES (Stanford Public Information REtrieval System), RIQS (Remote

----

[1]Gerald Jahoda. Information storage and retrieval systems for individual researchers. (New York: Wiley-Interscience, 1970)
A. H. Foskett. A guide to personal indexes using edge-notched, uniterm and peek-a-boo cards. 2nd ed. (London: Clive Bingley, c1970)
[2]Peter Leggate et al. "An on-line system for handling personal data bases on a PDP 11/20 minicomputer." Aslib Proceedings, 29(2):56-66, February 1977.
Benjamin Mittman and Lorraine Borman. Personalized data base systems. (Los Angeles: Melville Publishing Company, c1975.) 312 p.

Information Query System), and FAMULUS, to this problem will be discussed. Finally, an alternative system, based on a PL/1 program providing weighted-term searching, will be described.

I assume that the reader is familiar with the characteristics and use of conventional, coordinate, and edge-notched card systems, and therefore the section dealing with these is brief. Jahoda[3] describes their use in detail, and should be referred to if necessary for clarification.

Since I am assuming that a personal retrieval system will be primarily the work of one person, who has little time to spend on its development, this thesis will concentrate on the procedures used to store and subsequently to retrieve information, rather than on the recall and precision of the system. Techniques for improving system performance in these areas have been described previously[4]; since these involve greater depth of subject analysis and/or vocabulary control, they will therefore increase the time taken to add new references. The degree to which such measures are necessary should be determined by the user of the system, according to his own particular needs and standards.

I make no attempt to provide a quantitative analysis of the time taken to set up, alter, and use each system, since it will obviously vary from person to person, and will also depend on the depth of indexing done.

----

[3]Jahoda, Information storage and retrieval systems.
[4]Foskett, A guide to personal indexes. p.14-19.

I also do not discuss the physical arrangement of the items the researcher owns, since I feel that any logical arrangement will be satisfactory. Once the searcher finds that he does own a particular book, or has a copy of a specific article, he will probably know, at least roughly, where it is kept.

II. PERSONAL REFERENCE COLLECTIONS

In order to determine the retrieval needs of a researcher, it is first necessary to examine the characteristics and use of a personal reference collection. This will be done by means of a survey of the literature.

A. Characteristics

Little formal work has been done to identify the nature of personal document collections, perhaps because they vary too greatly for any meaningful conclusions to be drawn. The main study is that done by Burton[5], who found that such collections generally had well-defined patterns of input from a variety of sources, and tended to concentrate on primary literature. Additional information about personal reference collections can be found in the literature describing the ways in which researchers cope with the "information explosion". Examples of such articles are those by Cushing[6] and Hoff[7]. More recently, researchers have

----

[5]Hilary D. Burton. "Personal documentation methods and practices with analysis of their relation to formal bibliographic systems and theory." Microform Ph. D. thesis. University of California at Berkeley, 1972.
[6]Ralph Cushing. "A fresh look at improving personal filing systems." Chemical Engineering, 70(1):73-88, January 7, 1968.
[7]Wilbur J. Hoff. "A retrieval system for health education information." International Journal of Health Education, 9(2):87-93, June 1966.

described simple computer-aided systems. Bridges[8] describes a program which produces printed indexes, while Jameson[9] presents a batch retrieval system. It is interesting to note that few of the systems described in these articles have been developed in consultation with a librarian or information specialist. In fact, the authors emphasise the independence and adaptability of their systems.

However, it is possible to generalise somewhat about personal reference collections, and to make certain assumptions about their nature. First, such a collection is the work of only one person, and reflects his idiosyncrasies and interests. It includes a variety of material, some of which is actually owned by the researcher, with the rest available in the library, or perhaps in the collection of a colleague. The growth of the collection is erratic, depending on a number of factors, such as the stage of research and field of study. Obviously, when a new project is begun, the collection will grow rapidly as background material is gathered. When the research is well underway, the growth may slow down, since the researcher is not actively collecting information. Then, when a paper or report is being prepared, the collection will probably be used heavily (and possibly augmented), as references are

----

[8]Kent W. Bridges. "Automatic indexing of personal bibliographies." Bioscience, 20(2):94-97, January 15, 1970.
[9]David L. Jameson. "Information retrieval for the working scientist: a simple algorithm." Bioscience, 19(3):232-233, March 1969.

verified. As Yerke[10] points out, the relevance of a particular document will also depend on the stage of research, since new discoveries may lead to new hypotheses. These characteristics have implications for the type of retrieval system that will be suitable; these will be discussed in the next section.

B. Use of a personal reference collection

The two main studies of the ways in which personal reference collections are used were done by Jahoda, Hutchins, and Galford[11] and Burton[12]. Both found that the amount of information about an item included in such a file varied from person to person, and that where a retrieval system was in use, the access points also varied. However, the minimum information included was author, title, citation, and an indication of location; the most common access points used were subject, author, and source (including date). Searches, when done, had a range of complexities, but usually involved one to four concepts. Jahoda also studied the correspondence between title and subject, finding that title words alone could be used to locate relevant documents for over half the questions asked.

---

[10]Theodor B. Yerke. "Computer support of the researcher's own documentation." Datamation, 16(2):75-78, February 1970.
[11]Gerald Jahoda, R. D. Hutchins, and R. R. Galford. "Analysis of case histories of personal index use." American Documentation. Proceedings, 4:245-254, 1966.
idem. "Characteristics and use of personal indexes maintained by scientists at one university." American Documentation, 17(2):71-75, April 1966.
[12]Burton. "Personal documentation methods and practices."

The preceding discussion has implications for the design of a retrieval system for use by individual researchers. In fact, the standards first described by Krieger[13] still, for the most part, apply. According to Krieger, a personal retrieval system should:

1. find all information sought
2. allow for subjects not originally included
3. reduce duplication of entries to a minimum
4. involve as little coding as possible
5. avoid the use of complex sorting or punching machinery

A few more comments are necessary. First, because such a system will be established and maintained by only one person, the time required to set up, update, and use it should be kept to a minimum, yet the system should achieve the desired results. It should be able to accommodate a variety of information, including author, title, citation, and location. Searching for a number of concepts must be possible, and combinations of author, subject, and date must be searchable. As Yerke has pointed out, some documents may become obsolete or irrelevant, so, in addition, the system should allow for the removal of a record when it is no longer required.

To summarise, then, a retrieval system designed for use by an individual researcher should:

1. have a flexible input format

----

[13] K. A. Krieger. "A punched card system for chemical literature." Journal of Chemical Documentation, 26(3):163-166, March 1949.

2. be easy to use, and not involve extensive coding or checking of subjects
3. allow additional subject coverage
4. allow records no longer wanted to be removed
5. avoid duplication of entries

The above criteria will now be used to evaluate methods which have been suggested for use by individual researchers.

III. MANUAL SYSTEMS

The main card-based systems described for use by researchers are

1. the traditional author/title/subject system, similar to those found in most libraries
2. coordinate indexes, including both feature card indexes, such as the uniterm and optical coincidence systems, and edge-notched card systems

Each of these will be briefly described, then analysed to determine whether they meet the requirements discussed in the previous section.

A. Conventional Systems

Conventional systems are based on a set of subject headings, consisting of words or phrases. These may come from an existing list or thesaurus, such as the ERIC thesaurus, or the researcher may prefer to compile his own list. When an item is added to the collection, terms which describe its content are selected and a card made up for each access point. The cards are then kept in alphabetical order by subject heading. In addition, author/title entries may be made, with an entry for each author and also for the title. The amount of information on each card may vary, but there are two main forms.

In the first, complete information--i.e. author, title, citation, and location--appears only on the main author

card; the rest of the cards just give author and title. This means that subject searching is a two-step process. First relevant subject headings are consulted. Then, the information needed to find the items wanted is obtained from the author file.

The other form of the conventional card system involves repeating the full information about an item on each card relating to it. Searching is much simpler, but adding references is more repetitive, and also more time-consuming.

Obviously, such a system has both advantages and disadvantages for the researcher who wishes to use it. The main advantage is the flexibility of the subject headings. The researcher can be as specific as he likes when indexing items for retrieval, simply by modifying the phrase for the main subject. This can, however, also be a disadvantage, since it may be difficult to predict which aspects of an article will be of interest in the future. To re-assign subjects often would result in an extremely large file of subjects, and merely to discard cards no longer wanted would represent a waste of effort. Another disadvantage to the researcher is the difficulty of searching for a combination of subjects not represented by one subject heading. Jahoda[14] found that such a conventional subject index works well when there are three concepts or fewer per search, something not necessarily true in the case of the researcher, who is

----

[14]Gerald Jahoda. "A technique for determining index requirements." American Documentation, 15(2):82-85, April 1964.

likely to have a specific question to be answered. In addition, there is considerable duplication of effort, both in compiling the index, and in using it. First, information is repeated on each card, at least to some extent. Then, the researcher must spend time sorting and interfiling new cards into the existing index. While the actual amount of time spent will vary with the depth of subject analysis done and also with the growth rate of the collection, it is clear that much of the time spent in organising such a system is clerical in nature, and probably does not represent a valuable expenditure of a highly trained person's time. On the other hand, such a mindless activity may sometimes have therapeutic value particularly if the researcher is dealing with a stubborn problem.

The conventional card system, then, does not meet criterion number 5, since it involves a fair amount of duplication. If sufficient time and effort is spent in compiling such an index to a collection, the system is a good one, as libraries show. However, for the researcher who wants a simple and efficient system, it is definitely not suitable.

B. Coordinate Indexes

Coordinate indexes provide a method of searching for an almost unlimited number of subject terms, or descriptors, as they are commonly called. Such systems are found in two forms: the feature card system, in which there are two files, an accession file and a descriptor file, and the edge-notched card system, where there is only one file of cards.

Feature Card Systems

This type of index appears in two main forms, the "uniterm" system and the "optical coincidence" system. Since the two differ mainly in the nature of the card and equipment used, they will not be discussed separately, but will be described in a more general manner.

As I have stated above, feature card systems consist of two files: an accession file, in which each item included is given a unique number, usually assigned when it is first located; and a descriptor file, consisting of a card for each term used, indicating the accession numbers of items to which that term has been assigned. A typical accessions card will give author, title, citation, location, and list the descriptors assigned; it may also include brief comments about the item's usefulness. The cards must be kept in numerical

order. Foskett[15] suggests keeping it in a book, rather than on cards, since there is no reason to disarrange the entries. In the optical coincidence version, pre-numbered cards are used for the descriptor file, and the numbers of relevant documents punched, so that superimposing a number of descriptor cards will result in light shining through the locations of items given those descriptors. In the uniterm version, the card has ten columns; numbers of relevant items are written in the column of their final digit. Searching is done by comparing numbers until a match is found. In addition, cards can be kept for author and date, to provide access through these points.

Such a system has obvious advantages over the previously described traditional card system, since searching is much simpler. The number of terms used as descriptors is virtually unlimited, though a large number of terms results in a vast number of cards to be stored. Since the descriptor cards are kept in alphabetical order, and the accession file in numerical order, a great deal of filing may need to be done if the file is searched regularly. In addition, the number of documents that can be handled by such a system is limited by the size of the card, especially in the optical coincidence version. They are generally adequate for up to a thousand items, and special versions

----

[15] A. C. Foskett. A guide to personal indexes. p. 45.

requiring expensive precision drilling equipment can handle up to ten thousand documents. However, searching is a two-step procedure, and the need to keep two separate files again results in a great deal of clerical work, so that a busy researcher on his own may find it difficult to keep up.

Removing an item from the file is also difficult, particularly in the optical coincidence system, since the accession record and any occurrences of the number must be removed from the descriptor file.

Edge-notched Card Systems

Edge-notched cards are index cards with holes punched around their edges. A single hole or combination of holes stands for a particular term; when this term is used to index a document, the hole(s) is(are) punched. One card is used per document, and holes punched out as needed to indicate the descriptors, author, and source. Searching is done using a long needle; it is inserted in the location of the term of interest, and documents where this location has been punched will fall. Other terms are then searched on the fallen cards in order to do a complex search.

The holes can be assigned meaning in a number of ways. The simplest is to use one hole per subject, but this limits the number of terms to the number of holes on the card. In order to make more efficient use of the card, it is possible to use what is called indirect

coding, which means that two or more holes are used per descriptor. The most common is the random superimposed code[16], in which each term is assigned two random numbers; the holes for these numbers are punched when a document is assigned the term, and more than one descriptor is punched per field. Using a two hole code can greatly increase the number of terms that can be used, but it does make adding new records and searching more complicated, since the code book must be consulted each time to determine the code. It also increases the number of files that need to be kept, resulting in more clerical work to be done.

As many researchers have shown, however, an edge-notched system can be an effective means of providing access to a collection of references. It does, however, become time-consuming to search if there are more than a few hundred references, since only two hundred or so can be searched thoroughly at a time.

----

[16] Gerald Jahoda. Information storage and retrieval systems for individual researchers. p. 67.

C. Conclusions

Manual systems can provide an effective means of giving access to a researcher's collection, particularly if he has sufficient time to spend on the clerical details. However, there is considerable duplication of information. The need to consult several files before getting the results of the search is a major disadvantage, making searching a complicated procedure.

IV. DATA BASE MANAGEMENT SYSTEMS

In the last few years, data base management systems have become more and more common, partly because of the increasing sophistication of computers. They are capable of handling a variety of information, including bibliographic. Some examples are SPIRES (Stanford Public Information REtrieval System), RIQS (Remote Information Query System), and TDMS (Time-shared Data Management System). These are complex systems developed to meet a variety of needs, and a large number of users. In this thesis, I will discuss the use of SPIRES to handle bibliographic information, since it is the data base management system in use at the University of Alberta. I will not concentrate on specific examples of its use, but will instead describe the general procedures used to create, update, and search a data base, concentrating on the needs of the researcher.

A. Stanford Public Information REtrieval System

SPIRES is a data base management system developed at Stanford University in the early 1970's. It was originally designed as a system for library automation, but is now used for a variety of applications [17].

In order to use SPIRES, it is first necessary to define the file. The file definition process consists of determining the number of elements and characteristics of each element in a record. The language used for this is fairly complex, and it is unlikely that the average researcher would have the time (or inclination) to become familiar enough with it to define his own file. This means that he would have to hire a consultant to do so, and therefore some of the personal aspect is lost, since an outsider has become involved.

The researcher will probably have at least four elements in his file definition, all of them variable in both number of occurrences per record and also in length. The elements would be author, title, citation, and location. The citation could also indicate the type of material, that is, whether it is a book, periodical article, report, etc., in addition to the usual information. Alternatively, one could define the type of material and date of publication, asseparate elements, say, and SPIRES can easily allow this.

Setting up or adding records to SPIRES is a fairly complicated procedure. First, the information is placed in a

----

[17]G. R. Jackson. SPIRES/370 file definition. (Edmonton: The University of Alberta, 1977) p.5.

standard sequential line file, with each element identified in the form "ELEMENT-NAME=data;". Once the information is free of mistakes (and this alone can involve considerable effort), it is added to the data base using the SPIBILD processor. This is a relatively expensive procedure, since the data is compared with the file definition to make sure that it is in the correct form. In addition, indexes in the form of inverted files may be built for elements that are often searched. Indexes add to the cost of building the data base, but to use SPIRES efficiently, and therefore make searching cheaper, they are essential.

Searching is done using Boolean logic to combine terms; the cost of a search will depend on both the size of the subfile and on the number of terms used, as well as on whether an index has been built for the element(s) searched. Searching for unindexed elements is expensive, since the file must be searched sequentially to see if the wanted term or value occurs.

Items retrieved from a search are displayed or typed in the reverse order of input, that is, with the ones added most recently typed first. If he knows the SPIRES formatting language, the researcher can also vary output format according to his needs. This feature is useful if the output is to be used in a report or paper, but is probably not needed by the searcher who merely wants to know where a certain reference is located, or what he has collected on a certain subject.

From the preceding brief discussion of the characteristics of SPIRES, it can be seen that using a data base management system is a fairly awkward way of setting up a personal system. Creating a file to be searched is a complicated procedure, and storing the file in disk storage takes up a great deal of space, since not only the data, but also the file definition, formats, and any indexes which are created, must be stored. This will greatly increase the costs of maintaining the system, since as the data base grows in size, it will take up more and more disk space, and the owner pays for the amount of storage used. However, once the system has been set up, searching is simple and straightforward. One can alter existing records by transferring the complete record to a temporary line file, making changes using the operating system file editor, and then re-processing the record. An entry is removed by simply deleting it.

Overall, though, it is apparent that SPIRES in particular is not suitable for the researcher looking for a simple retrieval system to help him keep track of the literature in his field. It is interesting to note that none of the various applications described by Mittman and Borman[18] is a simple personal retrieval system; all are complex applications making use of the ability of the system to do much more than merely store and retrieve information,

----

[18]Benjamin Mittman and Lorraine Borman. Personalized data base systems. (Los Angeles: Melville Publishing Company, c1975)

and are in fact research data bases containing much more than simple bibliographic information. Data base management systems do provide an excellent means of providing access to a variety of information, and can analyse and plot data as well as retrieve it. They are also suitable for creating a data base for use by many people, since they can standardise data, and are also capable of providing instruction[19].

---

[19]ibid. p.3-33.

V. AN ALTERNATIVE RETRIEVAL METHOD

This chapter describes a program written in PL/1, based on a weighted-term algorithm previously described by Davis[20] The algorithm has been modified to allow the user to specify a maximum number of items retrieved, and also to allow him to search on a combination of elements. Although the program described here has been written in PL/1, it could be adapted to another language with little difficulty. PL/1 was chosen because of its built-in text-handling capabilities. The operating system used is MTS, since it is in use at the University of Alberta, but any modern operating system would be suitable. A program listing, sample data, and a sample search are given in the appendices to this thesis.

A full description of weighted-term searching is given in the article by Davis, and I will briefly summarise the main points. The searcher assigns arbitrary weights to each term used in the search. The weight of a particular record is then the sum of the weights of any search terms occurring in it; if it is greater than a threshold weight determined by the searcher, the information is printed out. Obviously, the search can be planned so that only certain combinations of terms will make the document weight greater than the threshold, and this allows a flexible search strategy.

The program has been written to make use of three

----

[20]Charles H. Davis. "A simple program for weighted term searching." Special Libraries, 63(9):381-384, September 1972.

fields, which can then be used for any data element. In the example I assign them to author, title, and citation/location. Data is stored in a standard line file, using the apostrophe as a delimiter; the end of the file is indicated by three 'ZZZ' lines . Each element can be up to 200 characters long, and therefore can be used for several "pieces" of data, if desired; for example, one could combine a title with added subject keywords. The use of three fields allows the user to distinguish between elements, but does not complicate input greatly. The program stores data in a two-dimensional array, A(j,k). The index j identifies a particular record and runs from 1 to 500, the maximum number of records; k runs from 1 to 3, corresponding to the three fields.

I am assuming that a personal collection larger than five hundred items can be subdivided in some way. The separation may be according to major subject, or it may be by something as arbitrary as date. In either case, the researcher will then have two files to search; he should be able to determine which would be the best to search. This assumption is supported by Engelbart[21], when he describes himself as using exactly this sort of subdivision to minimise the number of edge-notched cards to be searched at one time.

----

[21] Douglas C. Engelbart. "Special considerations of the individual as a user, generator, and retriever of information." American Documentation, 12(2):121-125, April 1966.

The use of three fields rather than one large one helps minimise the amount of data to be searched, since the program scans the file sequentially. This keeps the cost of the search down, while at the same time the input format is not complicated. Unlike SPIRES, no cpu time is used to process the data into inverted files. Storage is also kept to a minimum, since only the compiled program and data must be stored on disk.

Running the program is very simple, as can be seen from Appendix III. The searcher signs on, and issues the command "$run search+*pl1lib par=data=filename@u(200)", where "filename" is the name of the file in which the data is stored. The program first fills the data array from the external file in which the data is stored. Then, the searcher is asked to enter the maximum number of items he would like retrieved. I assume that only a few would be retrieved if the program is run in on-line mode, and that for a large search, batch mode would be used. Then, the searcher is asked to input the threshold weight and number of terms to be searched. Next, the searcher enters the type of search, using the numbers 1, 2, or 3 to indicate which element is to be searched. The search term is then entered, again using the apostrophe as a delimiter, along with its weight. This is repeated for each term. Then, the computer carries out the search, scanning the element for the occurrence of each search term. If a term occurs, the document weight is increased by the weight of the term, and

if the document weight is greater than the threshold weight, the record is printed. When either the maximum number of items has been retrieved, or the end of the data is reached, the searcher may continue with another search, or he can terminate the session by entering a 0(zero).

New records are added to the system by using the file editor capability of the operating system, and are immediately searchable. Changes are made in the same way. A record can be removed from the file simply by deleting the lines which apply to it.

One possible problem with this system is due to its simplicity. Since the file is searched sequentially, items are retrieved in the order in which they were input. A better method would be to arrange them in order of decreasing document weight, but this would complicate the program, and also increase the cost of searching, as one would have to store all retrieved records, prior to sorting by document weight. In principle, this could mean the addition of another 500 x 3 array.

It can be seen, though, that this program does provide an alternative to manual systems, by using the computer to minimise the more routine, clerical aspects of creating and maintaining a retrieval system, while at the same time avoiding the complexity of a data base management system.

VI. SUMMARY

Manual systems, data base management systems, and an alternative weighted-term system have been analysed in terms of the criteria identified in Chapter II. The following section roughly compares them in terms of cost and time needed to set up and use the system.

1. Cost of establishing

   uniterm < edge-notched cards < optical coincidence < weighted-term < SPIRES

   In the manual systems, the cost is mainly for materials, and will range from a few dollars for simple uniterm cards to several hundred or more for an optical coincidence system. The cost of the weighted-term system is for disk storage: a file of 200 records will cost approximately $50.00 per year. SPIRES records must be processed prior to storing, and this increases the cost to some $0.50 per record, or $200 annually for 200 items.

2. Time to set up

   weighted-term < edge-notched cards < uniterm, optical coincidence < SPIRES

3. Cost to search

   SPIRES < weighted-term (not applicable to manual systems)

   Searching for four terms in a 200 record file using the weighted-term system would cost about $1.50; the

same search would cost half that in SPIRES.

4. Time to search

SPIRES < weighted-term < optical coincidence < uniterm < edge-notched cards

From this, it is obvious that the cost and time involved in each system varies, and the researcher should choose the one that best suits his situation. Manual systems generally cost less, but are more time-consuming to search, while automated systems have a high initial cost, but are more easily searched. The weighted-term system described in this thesis presents an alternative to both manual and data base management systems, since it requires little time to set up and search, and costs are not prohibitive.

BIBLIOGRAPHY

Artandi, Susan. An introduction to computers in information science. 2nd ed. Metuchen: Scarecrow Press, 1972.

Bannur, B. B. and G. M. Purandare. "Personal documentation - a study." Annals of Library Science and Documentation, 14(4):182-186, December 1967.

Bourne, Charles P. Methods of information handling. New York: John Wiley & Sons, 1963.

Bridges, Kent W. "Automatic indexing of personal bibliographies." Bioscience, 20(2):94-97, January 15, 1970.

Burton, Hilary D. "Personal documentation methods and practices with analysis of their relation to formal bibliographic systems." Microform Ph.D. thesis, University of California, 1972. 1 microfilm, positive.

________. "Personal information systems: implications for libraries." Special Libraries, 64(1):7-11, January 1973.

________ and Theodor B. Yerke. "FAMULUS: a computer-based system for augmenting personal documentation efforts." In American Society for Information Science. Cooperating Information Societies. Proceedings of the annual meeting. 6:53-56, 1969.

Cushing, Ralph. "A fresh look at improving personal filing systems." Chemical Engineering, 70(1):73-88, January 7, 1968.

Davis, Charles H. Illustrative computer programming for

libraries: selected examples for information specialists. Westport: Greenwood Press, c1974.

________. An information retrieval primer. Ann Arbor, 1975. paper.

________. "A simple program for weighted-term searching." Special Libraries, 63(9):381-384, September 1972.

Engelbart, Douglas C. "Special considerations for the individual as a user, generator, and retriever of information." American Documentation, 12(2):121-125, April 1961.

Feinberg, Hilda. Title derivative indexing techniques. Metuchen: Scarecrow Press, 1973.

Foskett, A. C. A guide to personal indexes using edge-notched, uniterm and peek-a-boo cards. 2nd ed. London: Clive Bingley, c1970.

Heaps, Doreen M. and Paul Sorenson. "An on-line personal documentation system." In American Society for Information Science. Information transfer. Proceedings of the annual meeting, 5:201-207, 1968.

Hoff, Wilbur J. "A retrieval system for health education information." International Journal of Health Education, 9(2):87-93, June 1966.

Jackson, G. R. SPIRES/370 file defintion. Edmonton: The University of Alberta, 1977.

Jahoda, Gerald. Information storage and retrieval systems for individual researchers. New York: Wiley-Interscience, 1970.

________. "A technique for determining index requirements." American Documentation, 15(2):82-85, April 1964.

________, R. D. Hutchins, and R. R. Galford. "Analysis of case histories of personal index use." In American Documentation Institute. Proceedings of the annual meeting, 4:245-254, 1966.

________, ________, and ________. "Characteristics and use of personal indexes maintained by scientists and engineers in one university." American Documentation, 17(2):71-75, April 1966.

Jameson, David L. "Information retrieval for the working scientist: a simple algorithm." Bioscience, 19(3):232-233, March 1969.

Kent, Allen. Information analysis and retrieval. New York: Becker and Hayes, c1971.

Kochen, Manfred. Principles of information retrieval. Los Angeles: Melville Publishing Company, c1974.

Krieger, K. A. "A punched card system for chemical literature." Journal of Chemical Documentation, 26(3):163-166, March 1949.

Lancaster, F. Wilfred.Information retrieval systems: characteristics, testing, and evaluation. New York: John Wiley & Sons, c1968.

________ and E. G. Fayen. Information retrieval on-line. Los Angeles: Melville Publishing Company, c1973.

Leggate, Peter, B. M. Eaglestone, R. M. Jarman, M. M. Norgett, and A. P. Williams. "An on-line system for

handling personal data bases on a PDP 11/20 minicomputer." Aslib Proceedings, 29(2):56-66, February 1977.

Linn, William E., Jr., and Walter Reitman. "Referential communication in AUTONOTE, a personal information retrieval system." In Association for Computing Machinery. Proceedings of the annual conference. p. 67-82, 1971.

Loosjes, Th. P. On documentation of scientific literature. London: Butterworths, 1973.

McPherson, Arlean. "Computer indexing with S.I.S. II and FAMULUS." In American Society for Information Science. Proceedings: third annual meeting, Banff, October 3,4,5, 1971. Calgary: Information Systems, 1971. p. 121-128.

Marshall, K. E. "The evolution of a storage and retrieval system for indexes and annotated bibliographic references." In American Society for Information Science. Proceedings: third annual meeting, Banff, October 3,4,5, 1971. Calgary: Information Systems, 1971. p. 73-80.

Meadow, Charles T. The analysis of information systems. 2nd ed. Los Angeles: Melville Publishing Company, c1973.

Mittman, Benjamin and Lorraine Borman. Personalized data base systems. Los Angeles: Melville Publishing Company, 1975.

________ and Gilbert Krulee. "Development of a remote information management system - RIMS." In American

Society for Information Science. Cooperating information societies. Proceedings of the annual meeting. 6:199-206, 1969.

Norton, John H. "Setting up a personal information retrieval system." Management Review, 59(5):2-9, March 1970.

Vickery, Brian Campbell. Classification and indexing in science. 3rd ed. London: Butterworths, 1975.

Wallace, Everett M. "Experience with EDP support of individual's file maintenance." In American Documentation Institute. Proceedings of the annual meeting. 1:259-261, 1964.

Wilkinson, William M. "Indexing a personal reference file." Special Libraries, 50(1):16-18, January 1959.

Yerke, Theodor B. "Computer support of the researcher's own documentation." Datamation, 16(2):75-78, February 1970.

APPENDIX I: Program listing

```
SEARCH:    PROC OPTIONS(MAIN);
           DCL (A(500,3),B(3)) CHAR(200) VAR,
              WT(50) FIXED BIN,
              TERM(50) CHAR(100) VAR, TYPE(50) FIXED BIN,
              THRESH FIXED BIN, X FIXED BIN, DATA FILE,
              (NMAX,JUMP) FIXED BIN;
           ON CONVERSION GO TO ERROR;
           JUMP=0;
           NC=0;
  FILL:    /*FILL ARRAY*/
           GET FILE(DATA) LIST(B);
           IF B(1)='ZZZ' THEN GO TO START;
           NC=NC+1;
           DO I=1 TO 3;
              A(NC,I)=B(I);
           END;
           GO TO FILL;
 START:    /*OBTAIN SEARCH STRATEGY*/
           PUT SKIP EDIT('ENTER MAXIMUM NUMBER OF "HITS"')
              (A);
           PUT SKIP;
           IF JUMP=1 THEN PUT EDIT('TO END SEARCH, ',
              ENTER A 0 (ZERO)') (A,A);
           PUT SKIP;
           GET LIST(NMAX);
           IF NMAX=0 THEN GO TO QUIT;
```

```
PUT SKIP EDIT('ENTER THRESHOLD WEIGHT') (A);
PUT SKIP;
GET LIST(THRESH);
PUT SKIP EDIT('ENTER NUMBER OF TERMS') (A);
PUT SKIP;
GET LIST(X);
DO I=1 TO X;
   PUT SKIP EDIT('ENTER TYPE(',I,'), TERM(',I,'), ',
      'WEIGHT(',I,')') (A,F(2),A,F(2),A,A,F(2),A);
   PUT SKIP;
   GET LIST(TYPE(I),TERM(I),WT(I));
END;
PUT SKIP(3);
N=0;
DO IC=1 TO NC;
   DWT=0;
   DO I=1 TO X;
      IF INDEX(A(IC,TYPE(I)),TERM(I))¬=0 THEN
         DWT=DWT+WT(I);
   END;
   IF DWT>=THRESH & DWT¬=0 THEN GO TO PRINT;
   ELSE GO TO END1;
```

```
PRINT:         N=N+1;
               PUT SKIP(3) EDIT('DOCUMENT WEIGHT = ',DWT,'.')
                  (A,F(3),A);
               PUT SKIP;
               DO I=1 TO 3;
                  IF LENGTH(A(IC,I))<121 THEN PUT SKIP
                     EDIT(A(IC,I)) (A);
                  ELSE DO;
                     DO WHILE(LENGTH(A(IC,I))>120);
                        DO L=120 TO 1 BY -1 WHILE(SUBSTR(A(IC,I),L,1)
                           ¬=' ');
                        END;
                        PUT SKIP EDIT(SUBSTR(A(IC,I),1,L-1)) (A);
                        A(IC,I)=SUBSTR(A(IC,I),L+1);
                     END;
                     PUT SKIP EDIT(A(IC,I)) (A);
                  END;
               END;
               IF N=NMAX THEN GO TO END;
 END1:      END;
  END:      PUT SKIP(3) EDIT('NUMBER OF ITEMS FOUND = ',N,'.')
               (A,F(3),A);
            JUMP=1;
            GO TO START;
ERROR:      PUT SKIP(3) EDIT('ERROR.  A BLANK OR A PRIME IS PROBABLY ',
               'MISSING.') (A,A);
 QUIT:      END SEARCH;
```

APPENDIX II: Sample data

'ARIANDI, S.'
'AN INTRODUCTION TO COMPUTERS IN INFORMATION SCIENCE.'
'1972. Z 699 A72 1972'
'BANNUR, B.B. AND G.M. PURANDARE'
'PERSONAL DOCUMENTATION - A STUDY.'
'ANN. LIB. SCI. DOC. 14(4):182-286, DECEMBER 1967'
'BOURNE, C.H.'
'METHODS OF INFORMATION HANDLING.'
'1963. Z 699 B77'
'BRIDGES, K.'
'AUTOMATIC INDEXING OF PERSONAL BIBLIOGRAPHIES.'
'BIOSCIENCE, 20(2):94-97, JANUARY 15, 1970.'
'BURTON, H.D.'
'PERSONAL DOCUMENTATION METHODS AND PRACTICES WITH ANALYSIS OF THEIR RELATION TO FORMAL BIBLIOGRAPHIC SYSTEMS.'
'PH.D. THESIS. 1972'
'BURTON, H.D.'
'PERSONAL INFORMATION SYSTEMS: IMPLICATIONS FOR LIBRARIES.'
'SPEC. LIB., 64(1)7-11, JANUARY 1973.'
'CUSHING, R.'
'A FRESH LOOK AT IMPROVING PERSONAL FILING SYSTEMS.'
'CHEM. ENG., 70(1):73-88, JAN.7,1968.'
'DAVIS, C.H.'
'ILLUSTRATIVE COMPUTER PROGRAMMING FOR LIBRARIES: SELECTED EXAMPLES FOR INFORMATION SPECIALISTS.'
'1974'
'DAVIS, C.H.'
'AN INFORMATION RETRIEVAL PRIMER.'
'1975'
'DAVIS, C.H.'
'A SIMPLE PROGRAM FOR WEIGHTED-TERM SEARCHING.'
'SPEC. LIB., 63(9):381=384, SEPT. 1972.'
'ENGELBART, D.C.'
'SPECIAL CONSIDERATIONS OF THE INDIVIDUAL AS A USER, GENERATOR, AND RETRIEVER OF INFORMATION.'
'AM. DOC, 12(2):121-125, APRIL 1961.'
'FEINBERG, H.'
'TITLE DERIVATIVE INDEXING TECHNIQUES.'
'1973. Z695.92 F29 1973'
'FOSKETT, A.H.'
'A GUIDE TO PERSONAL INDEXES USING EDGE-NOTCHED, UNITERM, AND PEEK-A-BOO CARDS.'
'1970. Z 695.9 F743 1970'

'HEAPS, D.M. AND P. SORENSON.'
'AN ON-LINE PERSONAL DOCUMENTATION.'
'ASIS. PROCEEDINGS, VOL.5:201-207. 1968.'
'HOFF, W.J.'
'A RETRIEVAL SYSTEM FOR HEALTH EDUCATION INFORMATION.'
'INT. J. HEALTH ED., 9(2):87-93, JUNE 1966'
'JAHODA, G.'
'INFORMATION STORAGE AND RETRIEVAL SYSTEMS FOR INDIVIDUAL RESEARCHERS'
'1970'
'JAHODA, G.'
'A TECHNIQUE FOR DETERMINING INDEX REQUIREMENTS.'
'AM. DOC., 15(2):82-85, APR.1964.'
'JAHODA, G., E.D. HUTCHINS, AND R.R. GALFORD.'
'CHARACTERISTICS AND USE OF PERSONAL INDEXES MAINTAINED BY SCIENTISTS AND ENGINEERS IN ONE UNIVERSITY.'
'AM. DOC., 17(2):71-75, APR.1966.'
'JAHODA, G., R.D. HUTCHINS, AND R.R. GALFORD.'
'ANALYSIS OF CASE HISTORIES OF PERSONAL INDEX USE.'
'ADI. PROCEEDINGS, 4:245=254, 1966.'
'JAMESON, D.L.'
'INFORMATION RETRIEVAL FOR THE WORKING SCIENTIST: A SIMPLE ALGORITHM'
'BIOSCIENCE, 19(3):232-233, MAR.1969.'
'KENT, A.'
'INFORMATION ANALYSIS AND RETRIEVAL.'
'1971. Z 699 K365'
'LOOSJES, TH.P.'
'ON DOCUMENTATION OF SCIENTIFIC LITERATURE.'
'1973. Z 699 L86 1973'
'KOCHEN, M.'
'PRINCIPLES OF INFORMATION RETRIEVAL.'
'1974. Z 699 K764 1974'
'LANCASTER, F.W.'
'INFORMATION RETRIEVAL SYSTEMS: CHARACTERISTICS, TESTING, AND EVALUATION'
'1968. Z 699 L24'
'LANCASTER, F.W. AND E.G. FAYEN.'
'INFORMATION RETRIEVAL ON-LINE.'
'1973. Z 699 L238 1973'
'LEGGATE, P., B.M. EAGLESTONE, R.M. JARMAN, M.M. NORGETT, AND A.P. WILLIAMS.'
'AN ON-LINE SYSTEM FOR HANDLING PERSONAL DATA BASES ON A PDP 11/20 MINICOMPUTER'
'ASLIB PROCEEDINGS, 29(2):56-66, FEB.1977'

'MARSHALL, K.E.'
'THE EVOLUTION OF A STORAGE AND RETRIEVAL SYSTEM FOR INDEXED AND ANNOTATED BIBLIOGRAPHIC DATA BASES.'
'ASIS. WESCAN. PROCEEDINGS, 3RD, 1971, P. 73-80.'
'MCPHERSON, A.'
'COMPUTER INDEXING WITH S.I.S. II AND FAMULUS.'
'ASIS. WESCAN. PROCEEDINGS, 3RD, 1971, P. 121-128.'
'MEADOW, C.T.'
'THE ANALYSIS OF INFORMATION SYSTEMS.'
'1973. Z 699 M48 1973'
'MITTMAN, B. AND L. BORMAN.'
'PERSONALIZED DATA BASE SYSTEMS.'
'1975. Z 699 M67 1975'
'MITTMAN, B. AND G. KRULEE.'
'DEVELOPMENT OF A REMOTE INFORMATION MANAGEMENT SYSTEM - RIMS'
'ASIS. PROCEEDINGS. VOL.6:199-206, 1969.'
'NORTON, J.H.'
'SETTING UP A PERSONAL INFORMATION RETRIEVAL SYSTEM.'
'MANAGEMENT REV., 59(5):2-9, MAR.1970.'
'KREIGER, K.A.'
'A PUNCHED CARD SYSTEM FOR CHEMICAL LITERATURE.'
'J. CHEM. DOC., 26(3):163-166, MAR.1949.'
'VICKERY, B.C.'
'CLASSIFICATION AND INDEXING IN SCIENCE.'
'1975. Z 697 S5 V6 1975'
'WILKINSON, W.A.'
'INDEXING A PERSONAL REFERENCE FILE.'
'SPEC. LIB., 50(1):16-18, JAN. 1959'
'WALLACE, E.M.'
'EXPERIENCE WITH EDP SUPPORT OF INDIVIDUAL''S FILE MAINTENANCE.'
'ADI. PROCEEDINGS. VOL.1: 259-261, 1964.'
'YERKE, THEODOR B.'
'COMPUTER SUPPORT OF THE RESEARCHER''S OWN DOCUMENTATION.'
'DATAMATION, 16(2):75-78, FEB. 1970'
'ZZZ'
'ZZZ'
'ZZZ'
'ZZZ'
'ZZZ'
'ZZZ'

APPENDIX III: Sample search

```
#$RUN SEARCH+*=L1LIB PAR=DATA=BIBLIO...
#19:58:18


ENTER MAXIMUM NUMBER OF "HITS"

10


ENTER THRESHOLD WEIGHT
4

ENTER NUMBER OF TERMS
4

ENTER TYPE( 1), TERM( 1), WEIGHT( 1)
2 'PERSONAL' 2

ENTER TYPE( 2), TERM( 2), WEIGHT( 2)
2 'INDIVIDUAL' 3

ENTER TYPE( 3), TERM( 3), WEIGHT( 3)
2 'INDEX' 3

ENTER TYPE( 4), TERM( 4), WEIGHT( 4)
2 'RETRIEVAL' 2
```

DOCUMENT WEIGHT = 5.

BRIDGES, K.
AUTOMATIC INDEXING OF PERSONAL BIBLIOGRAPHIES.
BIOSCIENCE, 20(2):94-97, JANUARY 15, 1970.

DOCUMENT WEIGHT = 5.

FOSKETT, A.H.
A GUIDE TO PERSONAL INDEXES USING EDGE-NOTCHED, UNITERM, AND PEEK-A-BOO CARD
.
1970. Z 695.9 F743 1970

DOCUMENT WEIGHT = 5.

JAHODA, G.
INFORMATION STORAGE AND RETRIEVAL SYSTEMS FOR INDIVIDUAL RESEARCHERS
1970

DOCUMENT WEIGHT = 5.

JAHODA, G., R.D. HUTCHINS, AND R.R. GALFORD.
CHARACTERISTICS AND USE OF PERSONAL INDEXES MAINTAINED BY SCIENTISTS AND ENGIN
ERS IN ONE UNIVERSITY.
AM. DOC., 17(2):71-75, APR.1966.

DOCUMENT WEIGHT = 5.

JAHODA, G., R.D. HUTCHINS, AND R.R. GALFORD.
ANALYSIS OF CASE HISTORIES OF PERSONAL INDEX USE.
ADI. PROCEEDINGS, 4:245=25

```
MARSHALL, K.E.
THE EVOLUTION OF A STORAGE AND RETRIEVAL SYSTEM FOR   INDEXED AND ANNOTATED BI
LIOGRAPHIC DATA BASES.
ASIS. WESCAN.  PROCEEDINGS, 3RD, 1971, P. 73-80.


DOCUMENT WEIGHT =    4.

NORTON, J.H.
SETTING UP A PERSONAL INFORMATION RETRIEVAL SYSTEM.
MANAGEMENT REV., 59(5):2-9, MAR.1970.


DOCUMENT WEIGHT =    5.

WILKINSON, W.A.
INDEXING A PERSONAL REFERENCE FILE.
SPEC. LIB., 50(1):16-18, JAN. 1959


NUMBER OF ITEMS FOUND =    8.
ENTER MAXIMUM NUMBER OF "HITS"
TO END SEARCH, ENTER A 0(ZERO)
0

#20:00:39  T=.362  RC=0
#SIG $
#CCRN 19:57:15-20:00:46 TUE OCT 17/78
#    $.14 @ NON PRIME
```